

AD-A264 307



12

**CHEMICAL  
RESEARCH,  
DEVELOPMENT &  
ENGINEERING  
CENTER**

CRDEC-TR-420

**USE OF STATISTICS IN COMPUTATIONAL CHEMISTRY**

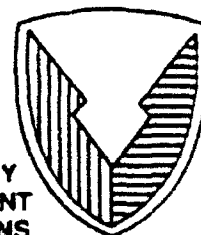
**Amnon Birenzvig  
Larry M. Sturdivan  
George R. Famini  
RESEARCH DIRECTORATE**

**P. N. Krishnan  
COPPIN STATE COLLEGE  
Baltimore, MD 21216**

**DTIC  
ELECTE  
MAY 14 1993  
S B D**

September 1992

Approved for public release; distribution is unlimited.



**U.S. ARMY  
ARMAMENT  
MUNITIONS  
CHEMICAL COMMAND**

Aberdeen Proving Ground, Maryland 21010-5423

**93 5 13 09 5**

**93-10809**



#### Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 1992 September	3. REPORT TYPE AND DATES COVERED Final, 90 Jun - 90 Aug	
4. TITLE AND SUBTITLE  Use of Statistics in Computational Chemistry			5. FUNDING NUMBERS  PR-1C162622A553	
6. AUTHOR(S) Birenzvege, Amnon, Sturdivan, Larry M., Famini, George R. (CRDEC), and Krishnan, P.N. (Coppin State College)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  CDR, CRDEC, ATTN: SMCCR-RSL, APG, MD 21010-5423  Coppin State College, Baltimore, MD 21216			8. PERFORMING ORGANIZATION REPORT NUMBER  CRDEC-TR-420	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT  Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  The physical and chemical properties of various molecules, calculated by three different semi-empirical methods (MNDO, AM1, and PM3) are compared to experimental values using rigorous statistical analysis. The results show that the best method for calculating physical/chemical properties depends on the property and that no one method is adequate to calculate all the physical properties. Furthermore, by employing appropriate statistical tools, we were able to determine the precision and accuracy that each method calculates the different physical/chemical properties. This will make it possible to use computational chemistry as a predictive tool with much better confidence.				
14. SUBJECT TERMS Accuracy Precision Chemical compounds			15. NUMBER OF PAGES 34	
Statistical methods Physical properties Computational chemistry			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT  UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE  UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT  UNCLASSIFIED	20. LIMITATION OF ABSTRACT  UL	

Blank

## PREFACE

The work described in this report was authorized under Project No. 1C162622A553, CB Defense/General Investigation. This work was started in June 1990 and completed in August 1990.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

Reproduction of this document in whole or in part is prohibited except with permission of the Commander, U.S. Army Chemical Research, Development and Engineering Center, ATTN: SMCCR-SPS-T, Aberdeen Proving Ground, MD 21010-5423. However, the Defense Technical Information Center and the National Technical Information Service are authorized to reproduce the document for U.S. Government purposes.

This report has been approved for release to the public.

RECEIVED 8

Accession For

NOV 1961	<input checked="" type="checkbox"/>
DEC 1961	<input type="checkbox"/>
JAN 1962	<input type="checkbox"/>

1962

1963

1964

1965

1966

1967

1968

1969

1970

1971

1972

1973

1974

1975

1976

1977

1978

1979

1980

1981

1982

1983

1984

1985

1986

1987

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019

2020

2021

2022

2023

2024

2025

2026

2027

2028

2029

2030

2031

2032

2033

2034

2035

2036

2037

2038

2039

2040

2041

2042

2043

2044

2045

2046

2047

2048

2049

2050

2051

2052

2053

2054

2055

2056

2057

2058

2059

2060

2061

2062

2063

2064

2065

2066

2067

2068

2069

2070

2071

2072

2073

2074

2075

2076

2077

2078

2079

2080

2081

2082

2083

2084

2085

2086

2087

2088

2089

2090

2091

2092

2093

2094

2095

2096

2097

2098

2099

2100

2101

2102

2103

2104

2105

2106

2107

2108

2109

2110

2111

2112

2113

2114

2115

2116

2117

2118

2119

2120

2121

2122

2123

2124

2125

2126

2127

2128

2129

2130

2131

2132

2133

2134

2135

2136

2137

2138

2139

2140

2141

2142

2143

2144

2145

2146

2147

2148

2149

2150

2151

2152

2153

2154

2155

2156

2157

2158

2159

2160

2161

2162

2163

2164

2165

2166

2167

2168

2169

2170

2171

2172

2173

2174

2175

2176

2177

2178

2179

2180

2181

2182

2183

2184

2185

2186

2187

2188

2189

2190

2191

2192

2193

2194

2195

2196

2197

2198

2199

2200

2201

2202

2203

2204

2205

2206

2207

2208

2209

2210

2211

2212

2213

2214

2215

2216

2217

2218

2219

2220

2221

2222

2223

2224

2225

2226

2227

2228

2229

2230

2231

2232

2233

2234

2235

2236

2237

2238

2239

2240

2241

2242

2243

2244

2245

2246

2247

2248

2249

2250

2251

2252

2253

2254

2255

2256

2257

2258

2259

2260

2261

2262

2263

2264

2265

2266

2267

2268

2269

2270

2271

2272

2273

2274

2275

2276

2277

2278

2279

2280

2281

2282

2283

2284

2285

2286

2287

2288

2289

2290

2291

2292

2293

2294

2295

2296

2297

2298

2299

2300

2301

2302

2303

2304

2305

2306

2307

2308

2309

2310

2311

2312

2313

2314

2315

2316

2317

2318

2319

2320

2321

2322

2323

2324

2325

2326

2327

2328

2329

2330

2331

2332

2333

2334

2335

2336

2337

2338

2339

2340

2341

2342

2343

2344

2345

2346

2347

2348

2349

2350

2351

2352

2353

2354

2355

2356

2357

2358

2359

2360

2361

2362

2363

2364

2365

2366

2367

2368

2369

2370

2371

2372

2373

2374

2375

2376

2377

2378

2379

2380

2381

2382

2383

2384

2385

2386

2387

2388

2389

2390

2391

2392

2393

2394

2395

2396

2397

2398

2399

2400

2401

2402

2403

2404

2405

2406

24

Blank

## CONTENTS

	Page
1. INTRODUCTION . . . . .	9
2. PRINCIPLES OF STATISTICAL METHODS USED . . . . .	9
3. RESULTS AND DISCUSSION . . . . .	13
3.1 Heat of Formation . . . . .	13
3.2 Ionization Potential . . . . .	13
3.3 Dipole Moment . . . . .	13
3.4 Bond Length . . . . .	14
3.5 Bond Angle . . . . .	14
4. CONCLUSIONS AND RECOMMENDATIONS . . . . .	14

## LIST OF FIGURES

1.	Comparison of Normal and Logistic Distribution. . . . .	11
2.	Test for Normal Distribution of Calculation Errors - Heat of Formation, PM3 . . . . .	15
3.	Test for Normal Distribution of Calculation Errors - Heat of Formation, MNDO . . . . .	16
4.	Test for Normal Distribution of Calculation Errors - Heat of Formation, AM1 . . . . .	16
5.	Test for Normal Distribution of Calculation Errors - Ionization Potential, MNDO . . . . .	17
6.	Test for Normal Distribution of Calculation Errors - Ionization Potential, PM3 . . . . .	17
7.	Test for Normal Distribution of Calculation Errors - Ionization Potential, AM1 . . . . .	18
8.	Test for Normal Distribution of Calculation Errors - Dipole Moment, PM3 . . . . .	19
9.	Test for Normal Distribution of Calculation Errors - Dipole Moment, MNDO . . . . .	20
10.	Test for Normal Distribution of Calculation Errors - Dipole Moment, AM1 . . . . .	21
11.	Test for Normal Distribution of Calculation Errors - Bond Length, PM3 . . . . .	22
12.	Test for Normal Distribution of Calculation Errors - Bond Length, MNDO . . . . .	23
13.	Test for Normal Distribution of Calculation Errors - Bond Length, AM1 . . . . .	24
14.	Test for Normal Distribution of Calculation Errors - Bond Length, PM3 - Extremes Excluded . . . . .	25
15.	Test for Normal Distribution of Calculation Errors - Bond Length, MNDO - Extremes Excluded . . . . .	26
16.	Test for Normal Distribution of Calculation Errors - Bond Length, AM1 - Extremes Excluded . . . . .	27
17.	Test for Normal Distribution of Calculation Errors - Bond Angle, PM3 . . . . .	28



18.	Test for Normal Distribution of Calculation Errors - Bond Angle, MNDO . . . . .	29
19.	Test for Normal Distribution of Calculation Errors - Bond Angle, AM1 . . . . .	30
20.	Test for Normal Distribution of Calculation Errors - Bond Angle, PM3 - Extremes Excluded . . . . .	31
21.	Test for Normal Distribution of Calculation Errors - Bond Angle, MNDO - Extremes Excluded . . . . .	32
22.	Test for Normal Distribution of Calculation Errors - Bond Angle, AM1 - Extremes Excluded . . . . .	33

#### LIST OF TABLES

1.	Compound with Largest Deviation of Calculated Error of Heat of Formation Using the PM3 Method . . . . .	15
2.	Compound with Largest Deviation of Calculated Error of Dipole Moment Using the PM3 Method . . . . .	19
3.	Compound with Largest Deviation of Calculated Error of Dipole Moment Using the MNDO Method . . . . .	20
4.	Compound with Largest Deviation of Calculated Error of Dipole Moment Using the AM1 Method . . . . .	21
5.	Compound with Largest Deviation of Calculated Error of Bond Length Using the PM3 Method . . . . .	22
6.	Compound with Largest Deviation of Calculated Error of Bond Length Using the MNDO Method . . . . .	23
7.	Compound with Largest Deviation of Calculated Error of Bond Length Using the AM1 Method . . . . .	24
8.	Compound with Largest Deviation of Calculated Error of Bond Angle Using the PM3 Method . . . . .	28
9.	Compound with Largest Deviation of Calculated Error of Bond Angle Using the MNDO Method . . . . .	29
10.	Compound with Largest Deviation of Calculated Error of Bond Angle Using the AM1 Method . . . . .	30
11.	Recommended Computational Methods for the Different Physical/Chemical Properties . . . . .	34

Blank

## USE OF STATISTICS IN COMPUTATIONAL CHEMISTRY

### 1. INTRODUCTION

With increased concern for the environment and the tightening of regulatory requirements, it is becoming increasingly more difficult and more expensive to conduct exploratory experiments in the chemical laboratory. With the significant advancements that have been made in recent years in the area of computational chemistry, it is now possible to calculate many physical and chemical properties using one of the commonly available computational methods. However, at this time, it is not possible to estimate how close the calculated property of a new compound is to the true value.

In a previous report, Birenzvice and co-workers<sup>1</sup> use statistical methods to compare the ability of the three most commonly used computational methods (MNDO, PM3, and AM1) to predict the heat of formation, dipole moment, polarizability and ionization potential of 12 vanillic-type molecules. The result of that study shows that by using proper statistical tools, we can predict the precision and accuracy of the computational method. In addition, Birenzvice and co-workers show that not one method is suited to calculate all the physical properties.

Stewart<sup>2</sup> published an up-to-date summary of computed heat of formation (775 molecules), geometries (bond length and bond angle - 209 molecules, including 174 bond angles and 372 bond lengths), dipole moment (125 molecules), and ionization potential (256 molecules). In this report, we extend the methods used before to compare how well the three computational methods can predict the various physical/chemical properties.

### 2. PRINCIPLES OF STATISTICAL METHODS USED

As previously stated, the purpose of this study was to determine the ability of each of the three semi-empirical methods

---

<sup>1</sup>Birenzvice, A., Sturdivan, L., Famini, G.R., Krishnan, P.N., and Morris, R.E., Predicting Polymer Properties by Computational Methods. 2: A Comparison of Semi-Empirical Methods, CRDEC-TR-361, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, September 1992, UNCLASSIFIED Report (A256 856).

<sup>2</sup>Stewart, James J.P., "Optimization of Parameters for Semi-empirical Methods, II. Applications," J. Comput. Chem. Vol 10, p 221 (1989).

to calculate the physical/chemical property in question. To enable us to determine the accuracy of the calculation (i.e., the standard deviation of the error) in a statistically meaningful way, we need to show that the calculation errors are symmetrically distributed about the real (experimental) values. This is conventionally done by showing that the data follow the normal distribution function. One way of showing that a data set is normally distributed is to order it in an ascending order and plot the data on a normal distribution graph paper. For example, suppose we take the weight of nine people ( $n=9$ ) selected at random. First, we sort them in ascending order; then we scale the linear Y axis so that all weights will fit. Finally, we plot the cumulative fraction on the probability axis versus the weight on the Y axis, letting the denominator of the fraction equal  $n+1$  (for symmetry). Thus, the lightest weight would be plotted versus 0.1 ( $1/n+1$ ), the next lightest versus 0.2, and so on until the heaviest would be plotted against 0.9 ( $n/n+1$ ). If the weights were normally distributed, the resulting nine points would fall on a straight line. Alternatively, we can calculate the normal score that is the expected value of the normal order statistic of an ordered sample of size  $n$ . In the statistical package MINITAB<sup>(tm)</sup>, the normal score is abbreviated N-score. Plotting the N-score against normally distributed data will result in the points falling about a straight line.

Calculating the N-score requires numerical solution of integral equations. Calculation of N-score is available in some statistical packages on minicomputers but is not available in commonly used software packages for microcomputers. To enable us to perform the analysis on a desk-top microcomputer, we need to find a distribution function that will closely resemble the normal distribution but will be easier to compute. The logistic distribution is such a distribution. Its straight line transform, which we will call the L-score, is obtainable in closed form and is simple to calculate.

Figure 1 shows a comparison of L-score and N-score. It was produced as follows: First, we calculated the N-score of an ordered set of numbers from 1 to 1000 using minitab<sup>(tm)</sup> on the VAX minicomputer. We then downloaded the data into a spread sheet on a desk-top PC and calculated the L-score according to the following equation:

$$L\text{-score}_{(i)} = \ln \frac{i}{n-i+1} \quad (1)$$

where  $i$  is the order of the item in the list, and  $n$  is the total number of items. The dashed line is the plot of L-score versus N-score. The solid line is a least-squares-fitted straight line

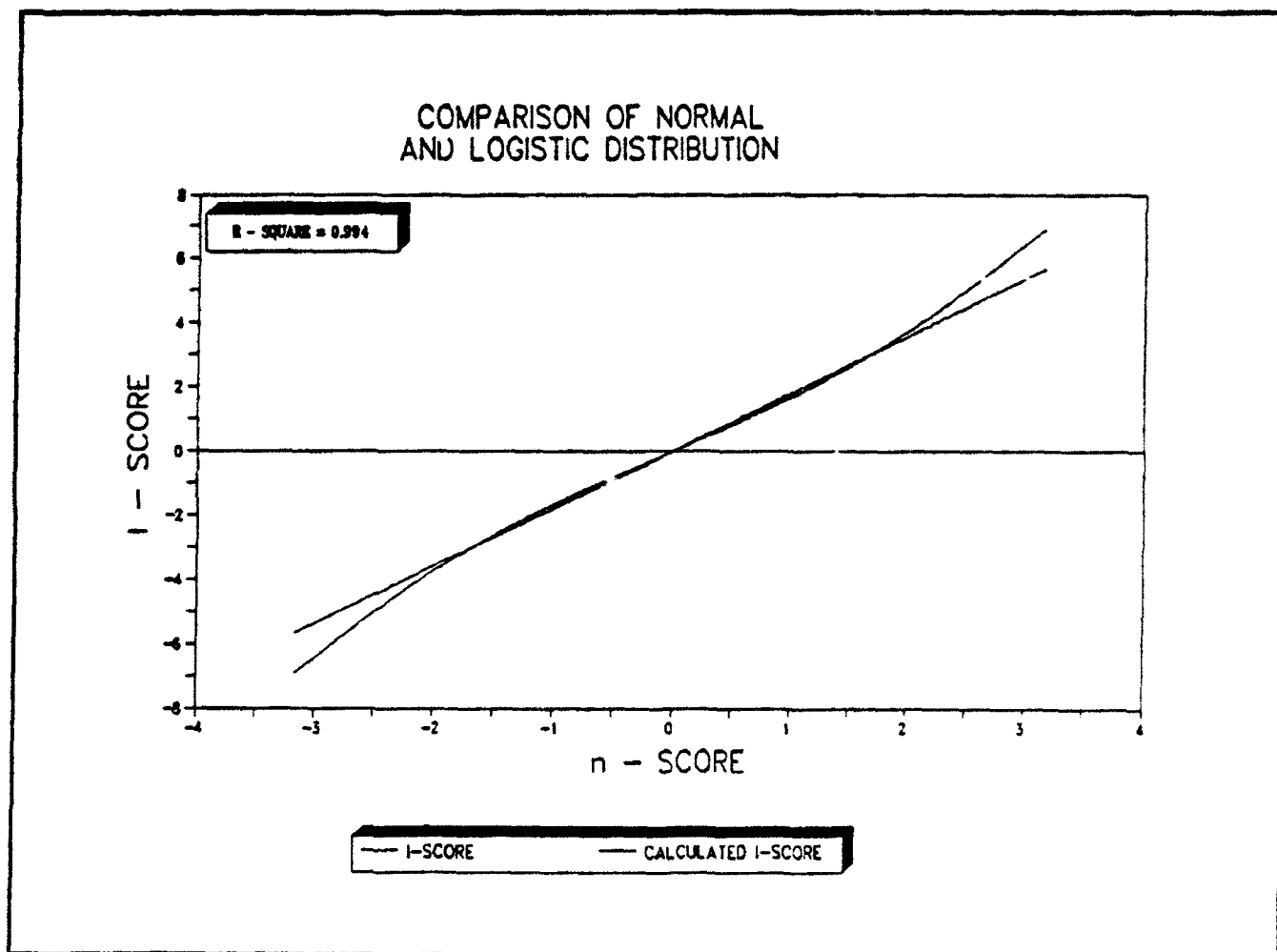


Figure 1. Comparison of Normal and Logistic Distribution

through the data. As can be seen, the two lines coincide except at the ends where the slightly heavier tails of the logistic distribution cause a bit of curvature away from the straight line. The correlation coefficient (R-squared) of the two measures is 0.994. When real data are plotted versus N-score and, separately, versus L-score, there is no visible difference in the "straightness," or lack of it, between the two plots.

For each molecule, the calculation errors (as reported by Stewart)<sup>2</sup> were plotted against L-score, and the correlation coefficient of the least squares regression line was determined. The plot was examined visually to determine any outliers and whether the fit will improve in a limited region. The average and standard deviation of the calculation errors were calculated in the region of symmetry as was the R-square for the least square regression line. To choose the best method for calculating the physical property in the region at a 95% confidence level, the following procedures should be followed:<sup>3</sup>

- For each of the three different methods, plot the difference between the calculated and experimental values of the property approximated versus its L-score. (Alternatively, plot the difference of the transformed data versus the L-score.)
- Determine a region (magnitude of calculated and experimental values) where the L-score plot forms a straight line, indicating a symmetrical "normal" distribution of errors. Transformed and untransformed data may have to be used for different regions (e.g., the data might be normally distributed in one region and lognormal in another).
- Calculate the R-square between the difference and the L-score for the appropriate region.
- Calculate the mean and standard deviation of the approximation error in the appropriate region.
- A method that has an R-square of 0.94 or larger is well approximated by the normal distribution. Among those that satisfy this criterion, choose the method that has the smallest standard deviation, unless that standard deviation is >2.28 times the size of the standard deviation of another method whose R-square is <0.94. In the latter case, choose the method with the smaller standard deviation regardless of the value of R-square.

---

<sup>3</sup>Mood, A.M., Graybill, F.A., and Boes, D.C., Introduction to the Theory of Statistics, 3rd ed., McGraw-Hill, New York, NY, 1974.

• Approximately 95% of the time, the experimental values will be in the range  $\langle (\text{calculated value} - \text{bias}) \pm 2\sigma \rangle$ .

### 3. RESULTS AND DISCUSSION

#### 3.1 Heat of Formation.

Figures 2-4 are plots of the differences between the calculated experimental heat of formation for PM3, MNDO, and AM1, respectively. In each case, the straight line represents the least squares regression line of the calculation errors versus L-score. As clearly seen only in the case of PM3, the R-square is larger than 0.94. The PM3 calculation method has an average bias of 0.44 Kcal/mole (i.e., the calculation errors are normally distributed about an error of 0.44 Kcal/mole) and a standard deviation of 12.3 Kcal/mole. Thus, the true heat of formation for a compound with no experimental value will be as follows:

$$\text{True Value} = (\text{Calculated Value} - 0.44) \pm 12.3 \times 2 \quad (2)$$

with 95% confidence.

Table 1 lists the seven compounds, which have the largest deviation (these compounds are marked in Figure 2). If these compounds are eliminated from the analysis, the precision (bias) and accuracy (standard deviation) of the method improve somewhat (0.37 Kcal/mole and 11.1 Kcal/mole, respectively).

#### 3.2 Ionization Potential.

Figures 5-7 show that the calculated ionization potential for all three methods are normally distributed about the experimental values. MNDO provides slightly higher accuracy (standard deviations of 0.7 ev versus 0.73 ev for PM3 and AM1, respectively) and should be the preferred method even though it has higher bias (0.68 ev versus 0.12 ev and 0.40 ev for PM3 and AM1, respectively).

#### 3.3 Dipole Moment.

The correlation coefficient of the calculation errors of the dipole moment versus L-score for PM3 (Figure 8) is only 0.92; thus, this method is not suitable for calculating the dipole moment. The correlation coefficient for both MNDO (Figure 9) and AM1 (Figure 10) is 0.96. The standard deviation for AM1 is smaller (0.49 deby) than the standard deviation for MNDO (0.63 deby). Thus, AM1 should be the preferred method for calculating dipole moment. The table under each figure (Tables 2-4) lists the compounds with the largest deviation from the regression line.

### 3.4 Bond Length.

As can be seen (Figures 11-13), none of the three methods evaluated in this study produces a data base of bond length, which is normally distributed about the true (experimental) values. In each case, there were few clear outliers, which are listed in Tables 5-7 (nine compounds for PM3, five compounds for MNDO, and seven compounds for AM1). When these outliers were removed from considerations, the correlation coefficient increased considerably (Figures 14-16). However, in the case of PM3 and MNDO, the correlation coefficient still remained below 0.94. In the case of AM1 when points 1-5 and point 8 were removed (these points show the largest deviation from the regression line), the correlation coefficient increased to 0.94.

### 3.5 Bond Angle.

Of the three methods evaluated (Figures 17-19), it appears that only AM1 shows a tendency for symmetry (Figure 19) even though there are few clear outliers. After removing the most extreme outliers from consideration in PM3 and MNDO (Tables 8-9), the correlation coefficient in PM3 increases to 0.92 (Figure 20) - not enough to show that the data are normally distributed. The correlation coefficient for MNDO remains very low (Figure 21). For AM1, once the six compounds listed in Table 10 are removed from consideration, the correlation coefficient between the computation errors and their corresponding L-score exceeds the required 0.94 (Figure 22).

## 4. CONCLUSIONS AND RECOMMENDATIONS

By using appropriate statistical tools, we show that semi-empirical computational methods can be used as screening methods to predict different physical or chemical compounds with a high degree of confidence. Furthermore, we show that not one of the commonly used computational methods is adequate for computing all of the properties.

For each of the properties studied, the computation error can be divided by a systematic error or bias (the precision of the calculation) and a random error (the accuracy of the calculation or standard deviation). Table 11 lists the different properties, the recommended computational method, and its bias and standard deviation. When calculating the property of an unknown compound, the true value can be evaluated as follows (with 95% confidence):

$$\text{True Value} = (\text{Computed Value} - \text{Bias}) \pm 2\sigma \quad (3)$$



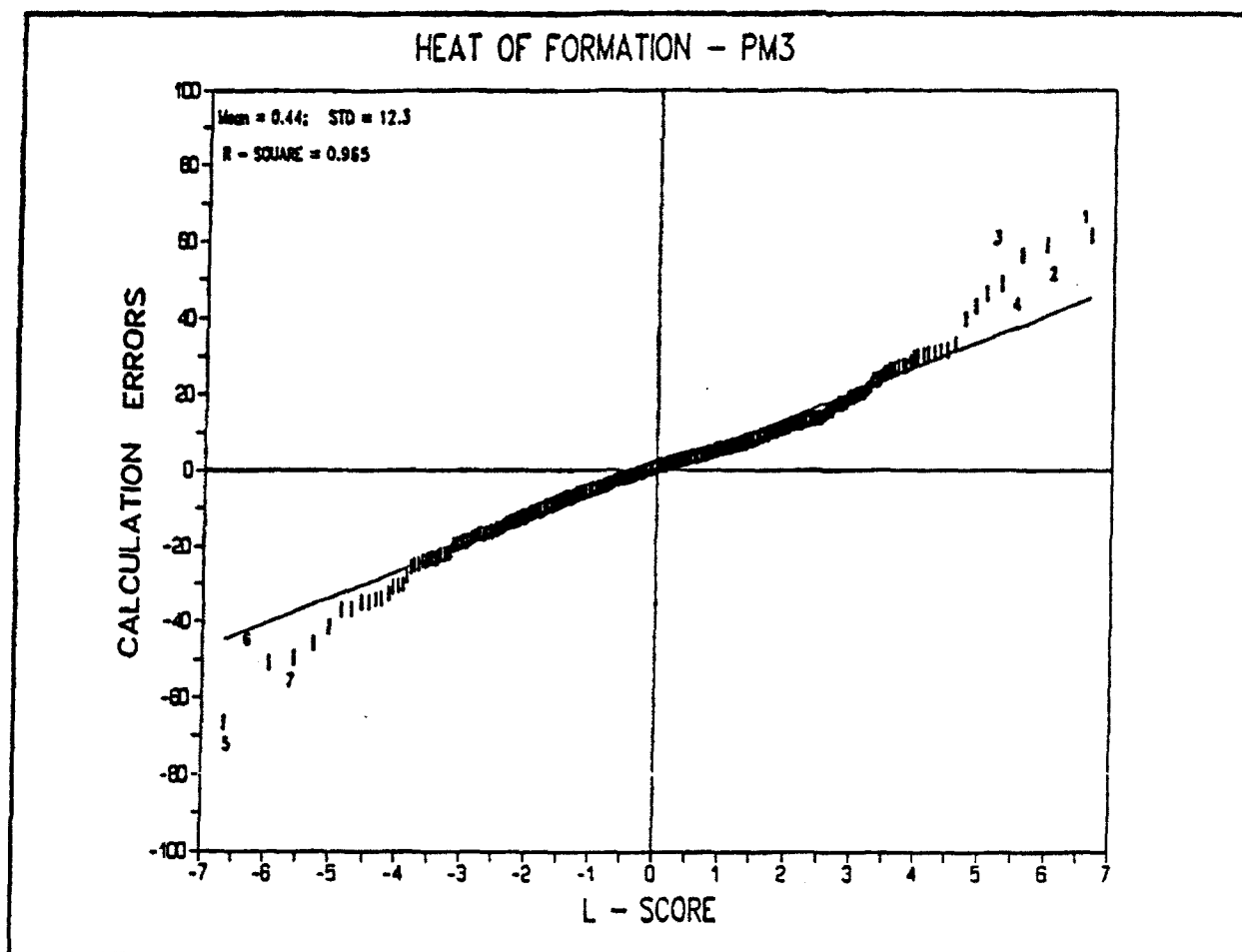


Figure 2. Test for Normal Distribution of Calculation Errors - Heat of Formation, PM3

Table 1. Compound with Largest Deviation of Calculated Error of Heat of Formation Using the PM3 Method

#	CHEMICAL	EXPERIMENTAL	CALC. ERRORS
1	Al <sup>(+)</sup>	218.1	61.7
2	PSBr <sub>3</sub>	-67.2	59.4
3	AlOF <sub>2</sub>	-265	56.5
4	SiOI <sub>2</sub>	-99.4	49.2
5	COI	63.5	-66.4
6	AlOH <sup>(-)</sup>	-55	-50.7
7	AlN	125	-49.5

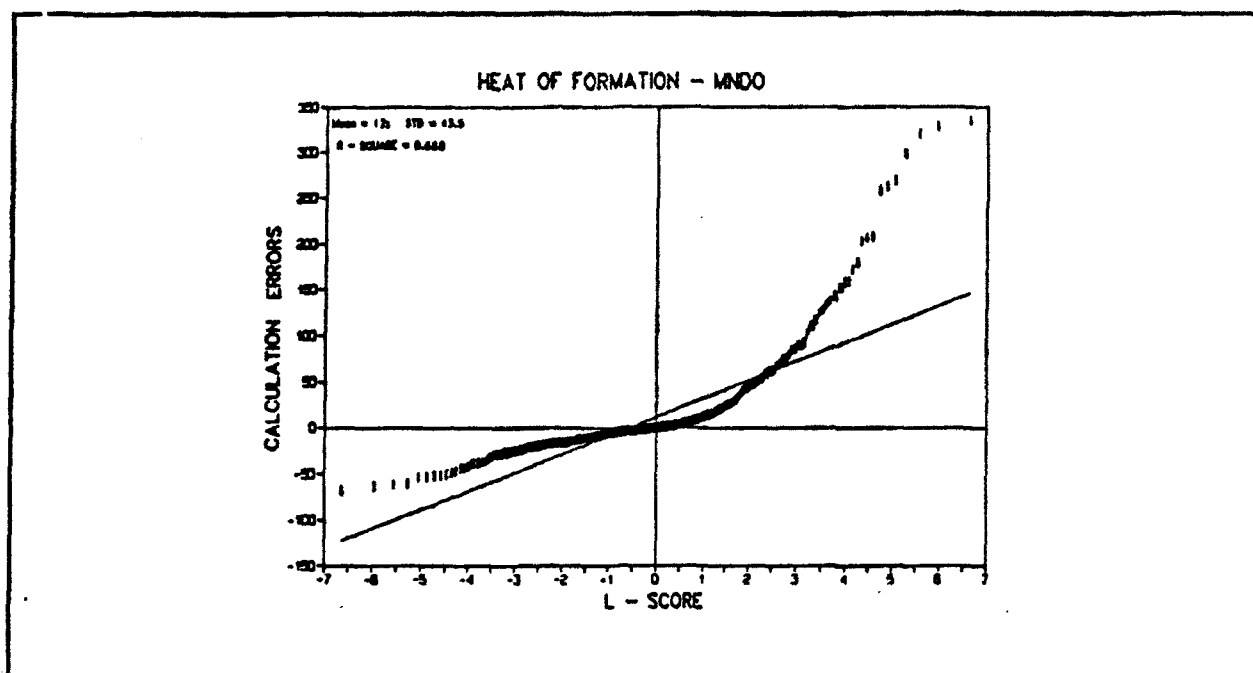


Figure 3. Test for Normal Distribution of Calculation Errors - Heat of Formation, MNDO

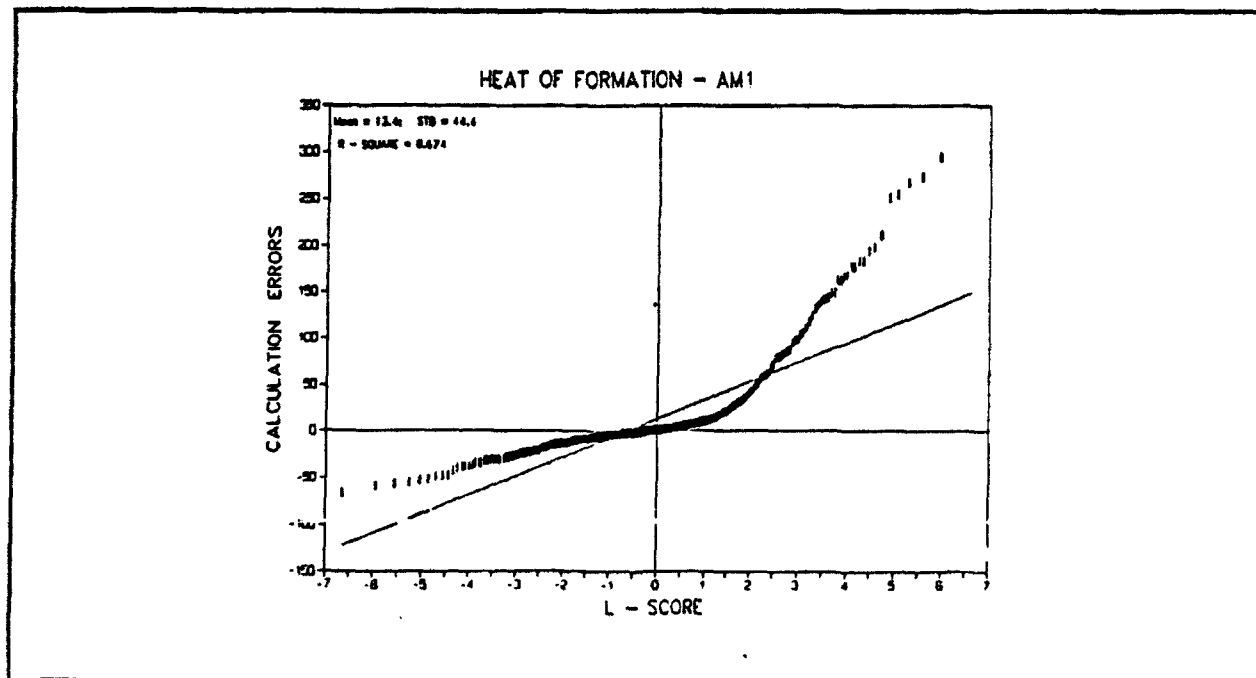


Figure 4. Test for Normal Distribution of Calculation Errors - Heat of Formation, AM1

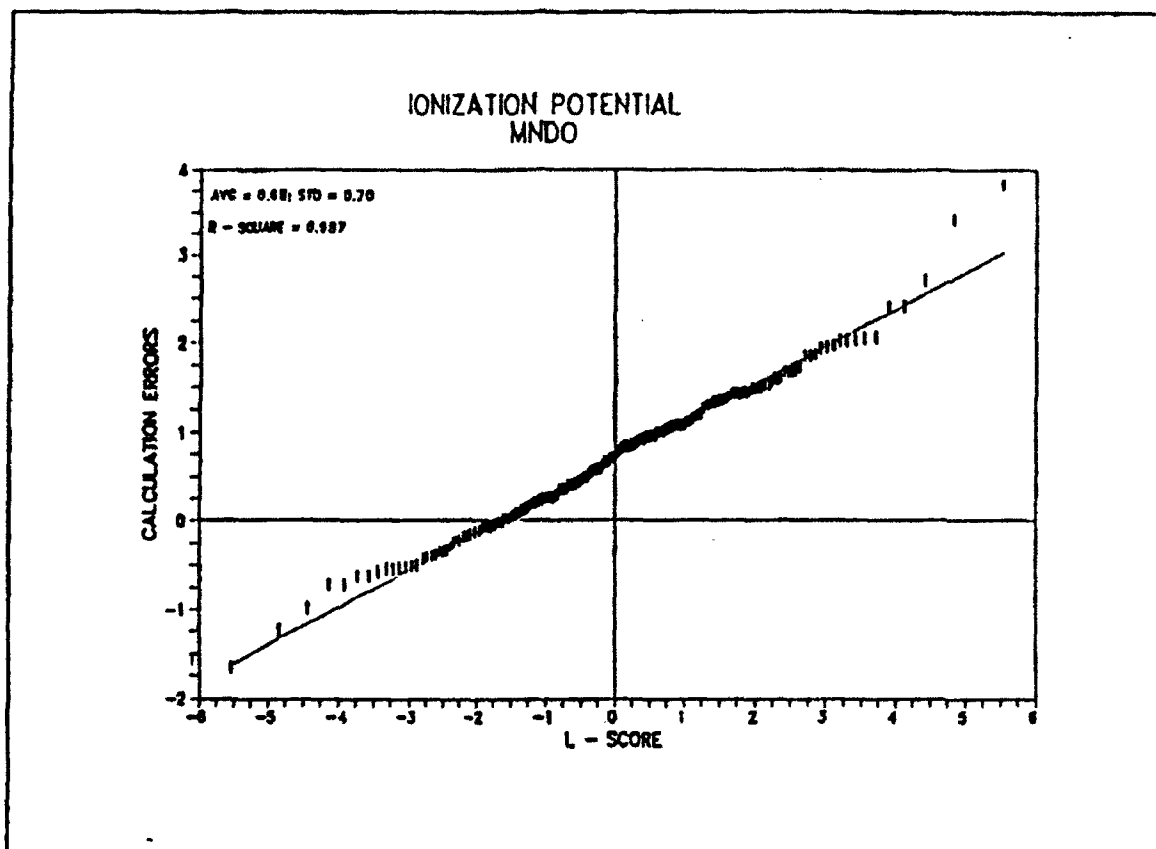


Figure 5. Test for Normal Distribution of Calculation Errors  
- Ionization Potential, MNDO

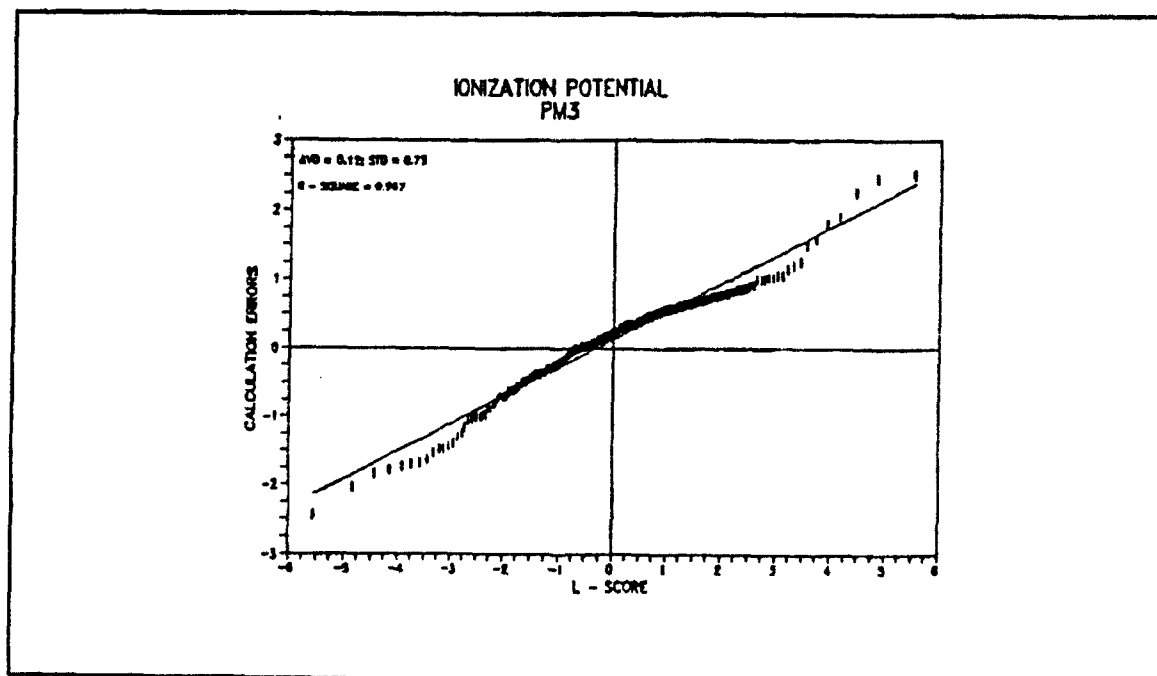


Figure 6. Test for Normal Distribution of Calculation Errors  
- Ionization Potential, PM3

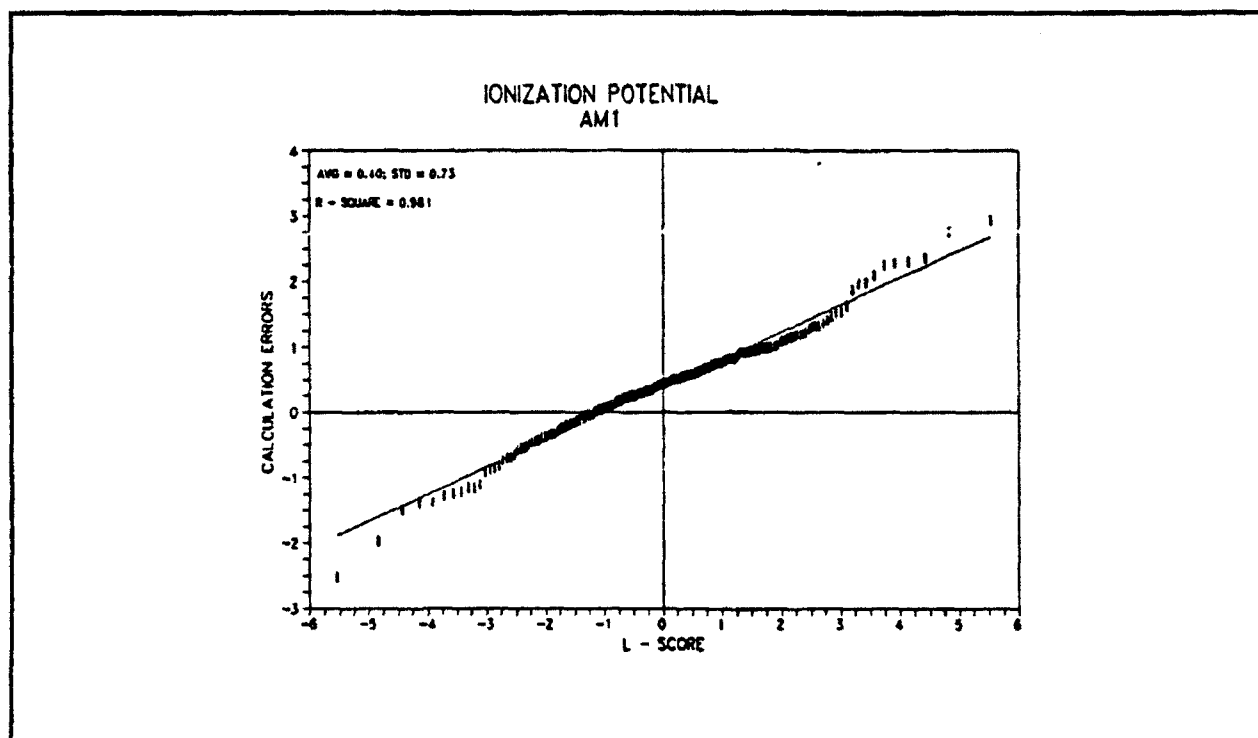


Figure 7. Test for Normal Distribution of Calculation Errors  
- Ionization Potential, AM1

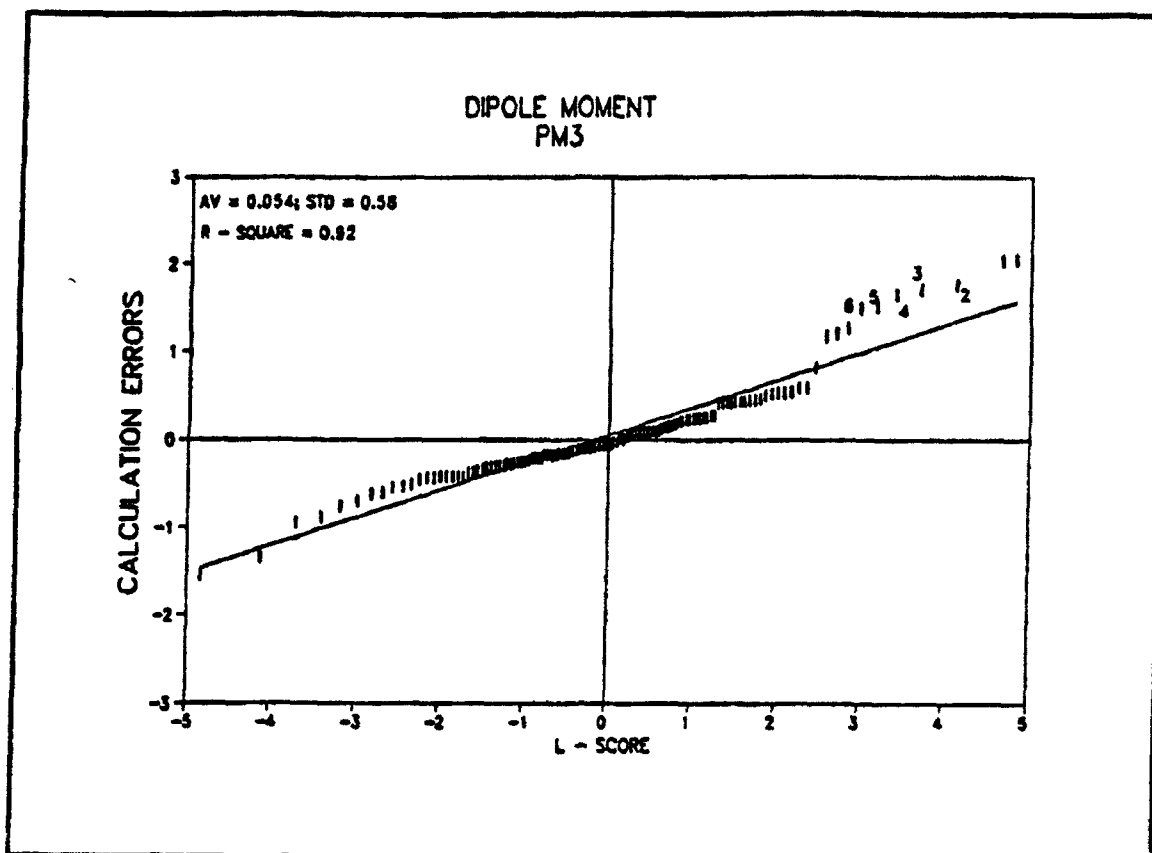


Figure 8. Test for Normal Distribution of Calculation Errors  
- Dipole Moment, PM3

Table 2. Compound with Largest Deviation of Calculated Error  
of Dipole Moment Using the PM3 Method

#	CHEMICAL	EXPERIMENTAL	CALC. ERRORS
1	SO <sub>2</sub>	1.57	2.06
2	AlF	1.53	1.76
3	BrO	1.61	1.71
4	H <sub>2</sub> SiBr <sub>2</sub>	1.43	1.66
5	HSiCl <sub>3</sub>	0.86	1.52
6	H <sub>2</sub> SiCl <sub>2</sub>	1.18	1.5

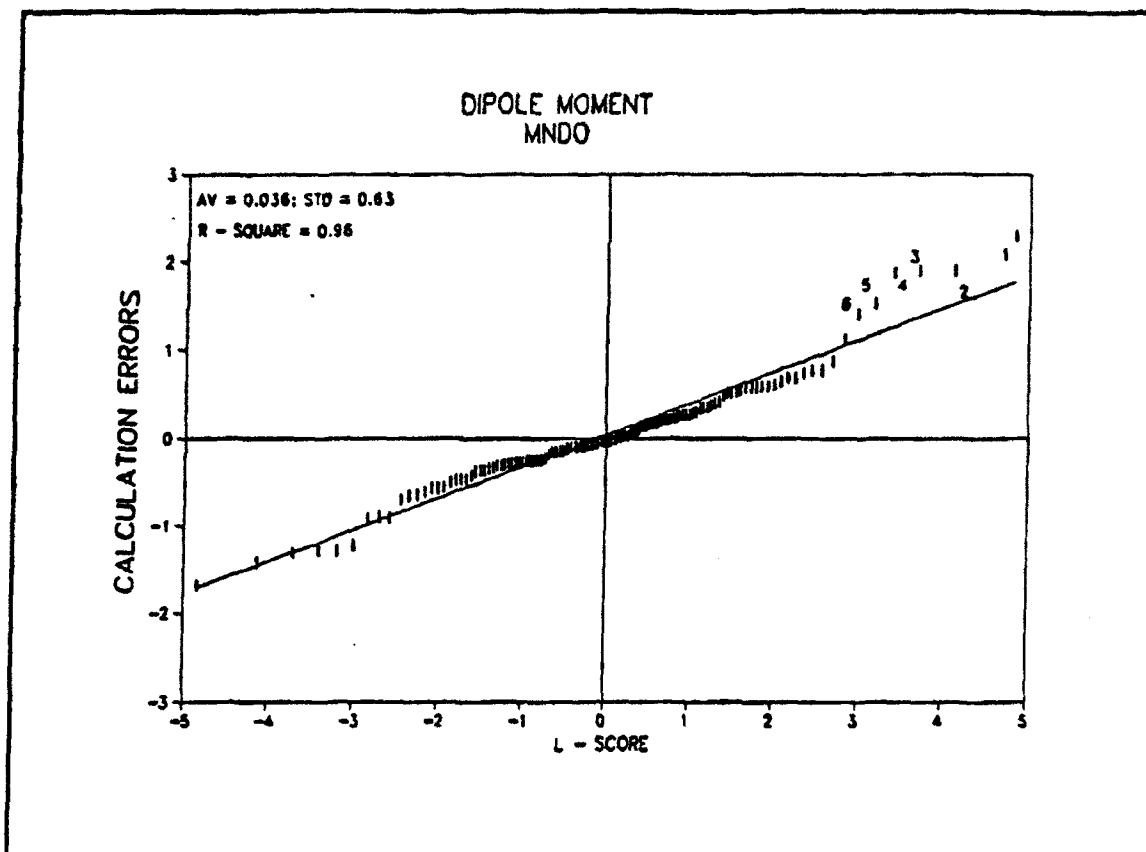


Figure 9. Test for Normal Distribution of Calculation Errors  
- Dipole Moment, MNDO

Table 3. Compound with Largest Deviation of Calculated Error  
of Dipole Moment Using the MNDO Method

#	CHEMICAL	EXPERIMENTAL	CALC. ERRORS
1	H <sub>2</sub> SiCl <sub>2</sub>	1.18	2.29
2	H <sub>2</sub> SiBr <sub>2</sub>	1.43	1.9
3	SO <sub>2</sub>	1.57	1.9
4	HSiCl <sub>3</sub>	0.86	1.87
5	HSiF <sub>3</sub>	1.27	1.53
6	PF <sub>3</sub>	1.03	1.22

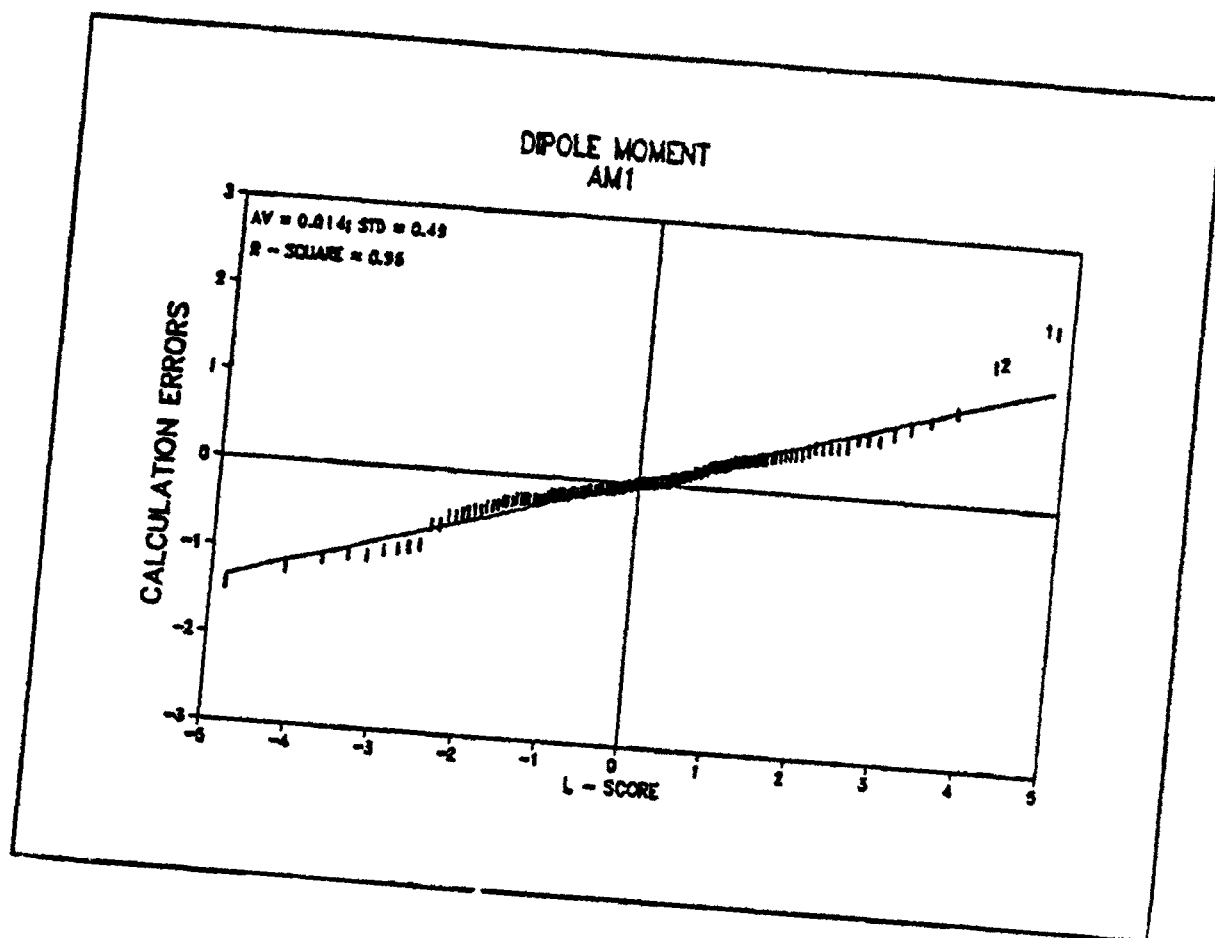


Figure 10. Test for Normal Distribution of Calculation Errors  
- Dipole Moment, AM1

Table 4. Compound with Largest Deviation of Calculated Error  
of Dipole Moment Using the AM1 Method

#	CHEMICAL	EXPERIMENTAL	CALC. ERRORS
1	SO <sub>2</sub>	1.57	2.06
2	H <sub>3</sub> P	0.58	1.61

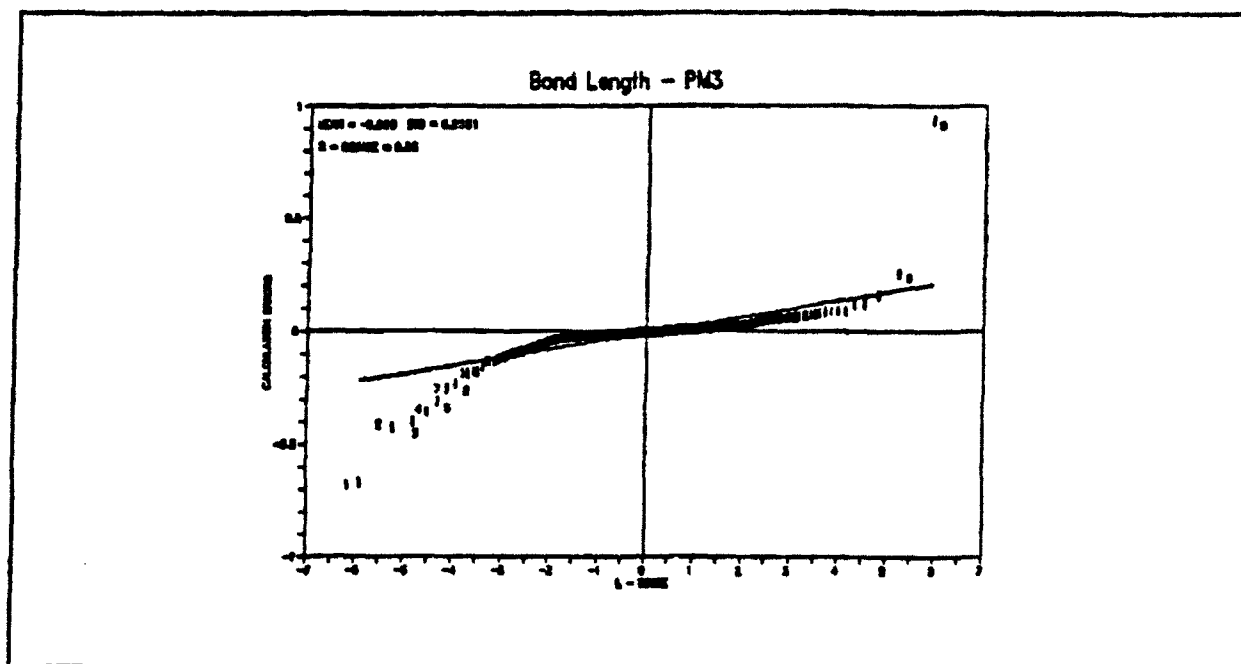


Figure 11. Test for Normal Distribution of Calculation Errors  
- Bond Length, PM3

Table 5. Compound with Largest Deviation of Calculated Error  
of Bond Length Using the PM3 Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	N <sub>2</sub> O <sub>3</sub>	N-N	2.08	-0.671
2	H <sub>3</sub> SiI	Si-I	2.012	-0.425
3	AlBr <sub>3</sub>	Al-Br	2.27	-0.395
4	SiBr <sub>4</sub>	Si-Br	2.15	-0.354
5	F <sub>3</sub> I	I-F(ax)	1.76	0.938
6	H <sub>3</sub> SiBr	Si-Br	2.21	-0.309
7	NOBr	Br-N	2.14	-0.252
8	H <sub>2</sub> O <sub>2</sub> (H <sub>2</sub> O Dimer)	O-O	3	-0.231
9	H <sub>2</sub> SO <sub>4</sub>	S-O	1.42	0.248



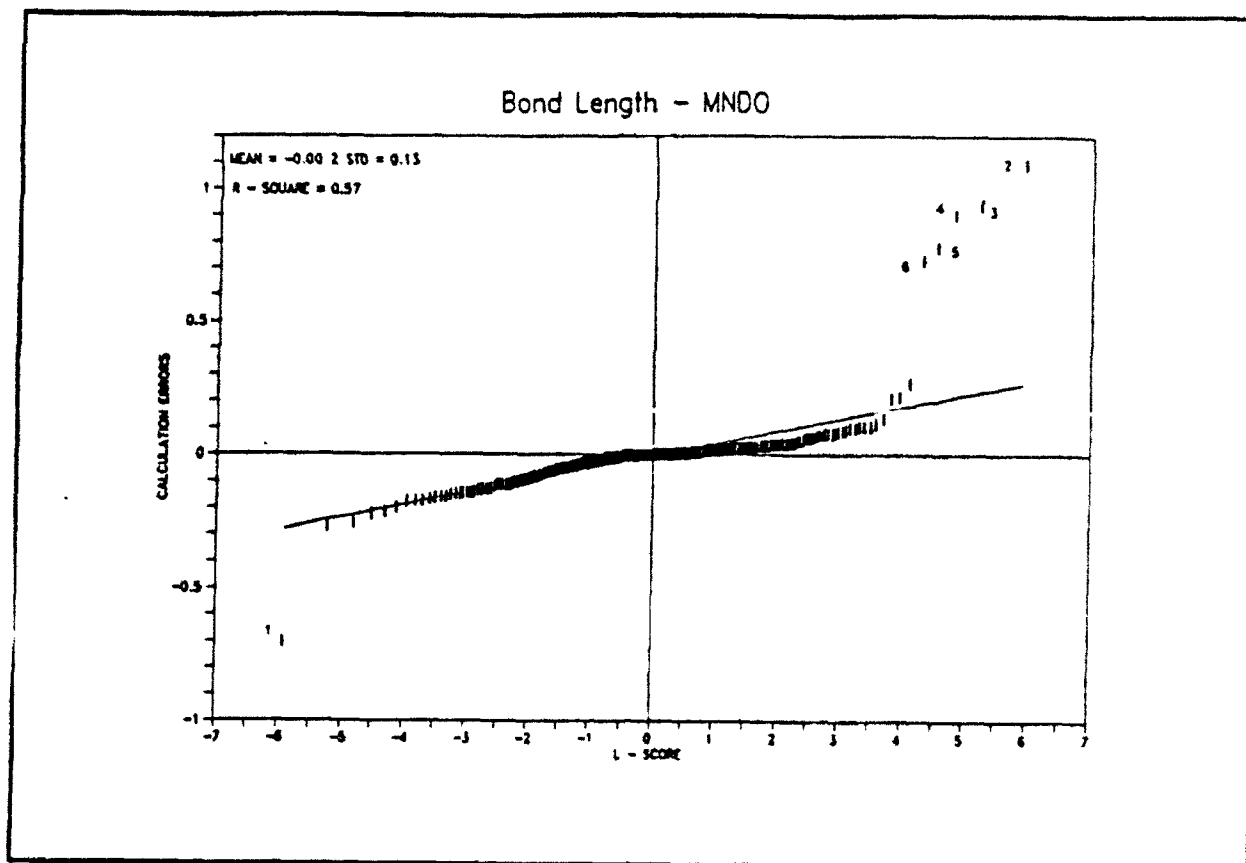


Figure 12. Test for Normal Distribution of Calculation Errors  
- Bond Length, MNDO

Table 6. Compound with Largest Deviation of Calculated Error  
of Bond Length Using the MNDO Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	N <sub>2</sub> O <sub>2</sub>	N-N	2.08	-0.706
2	H <sub>2</sub> F <sub>2</sub> (HF dimer)	H'-F	1.87	1.093
3	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub> (salicylaldehyde)	N(14)-H(17)	1.834	0.94
4	H <sub>2</sub> O <sub>2</sub> (H <sub>2</sub> O dimer)	O-O	3.0	0.905
5	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub> (salicylaldehyde)	O(10)-N(14)	2.626	0.781
6	F <sub>2</sub> I	I-F (ax)	1.76	0.731

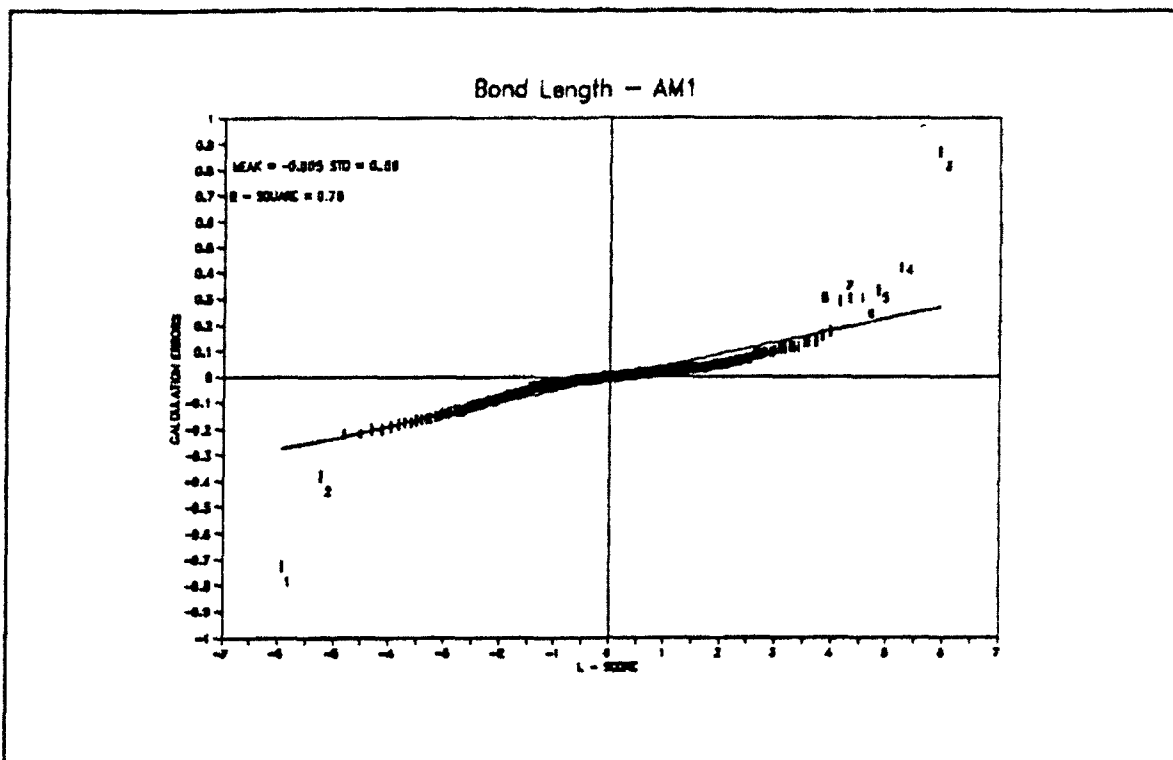


Figure 13. Test for Normal Distribution of Calculation Errors  
- Bond Length, AM1

Table 7. Compound with Largest Deviation of Calculated Error  
of Bond Length Using the AM1 Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	N <sub>2</sub> O <sub>2</sub>	N-N	2.08	-0.728
2	H <sub>2</sub> O <sub>2</sub> (H <sub>2</sub> O dimer)	O-O	3.00	-0.383
3	F <sub>2</sub> I	I-F (ax)	1.76	0.866
4	H <sub>2</sub> F <sub>2</sub> (HF dimer)	H'-F	1.87	0.417
5	O <sub>2</sub> FC1	Cl-O	1.46	0.328
6	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub> (Salicylaldehyde)	N(14)-H(17)	1.834	0.302
7	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub> (Salicylaldehyde)	O(10)-N(14)	2.684	0.298
8	H <sub>2</sub> SO <sub>4</sub>	S=O	1.42	0.287

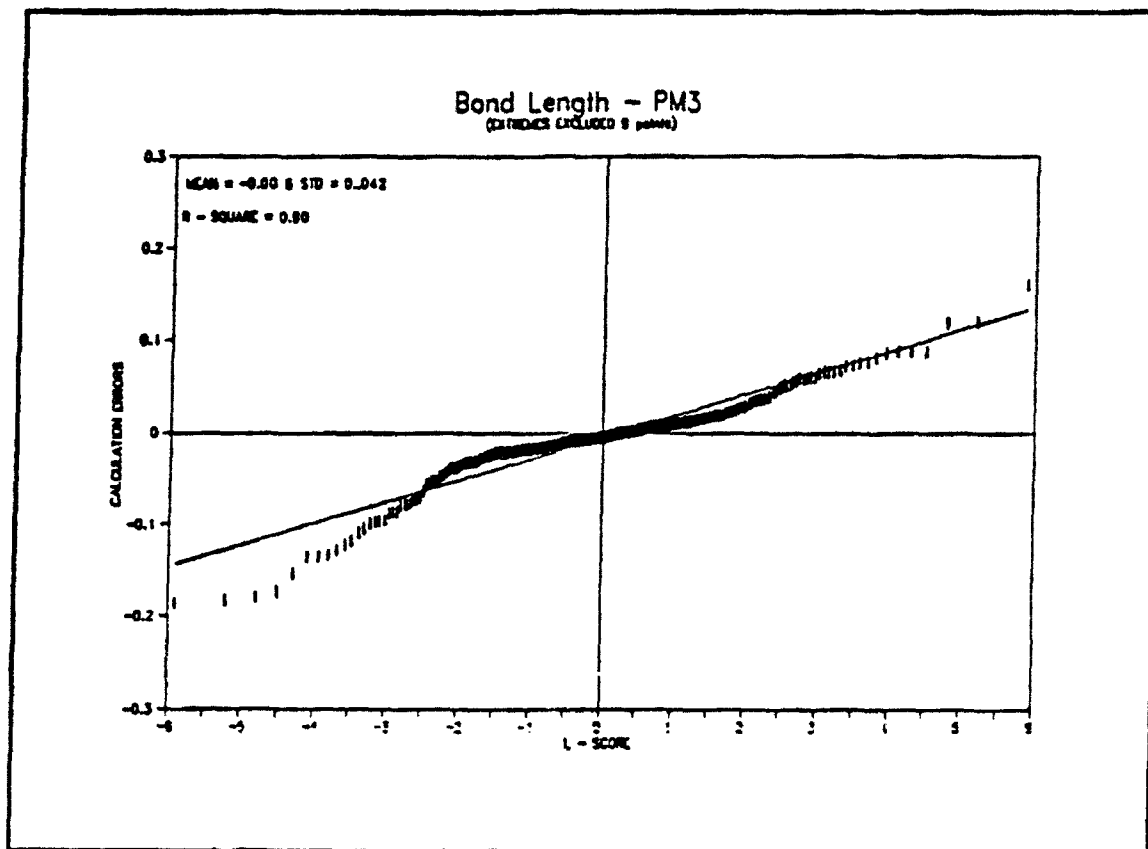


Figure 14. Test for Normal Distribution of Calculation Errors  
- Bond Length, PM3 - Extremes Excluded

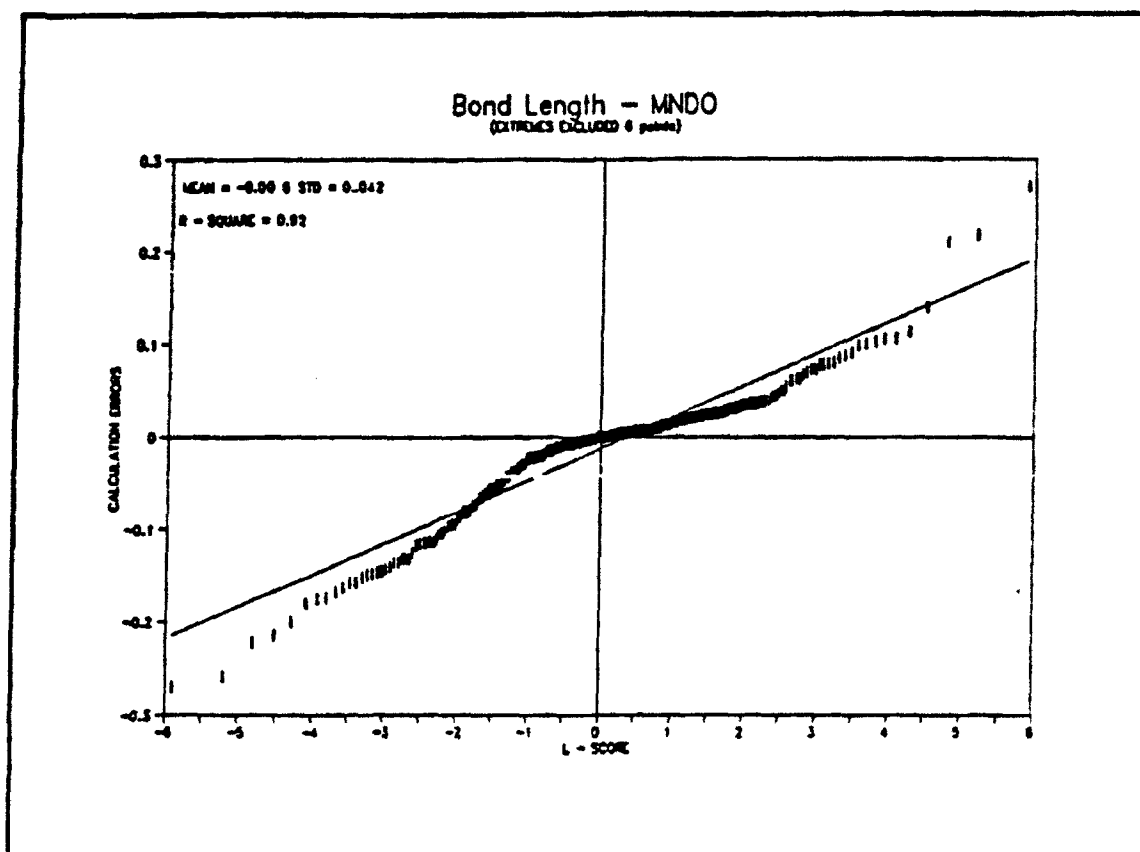


Figure 15. Test for Normal Distribution of Calculation Errors  
- Bond Length, MNDO - Extremes Excluded

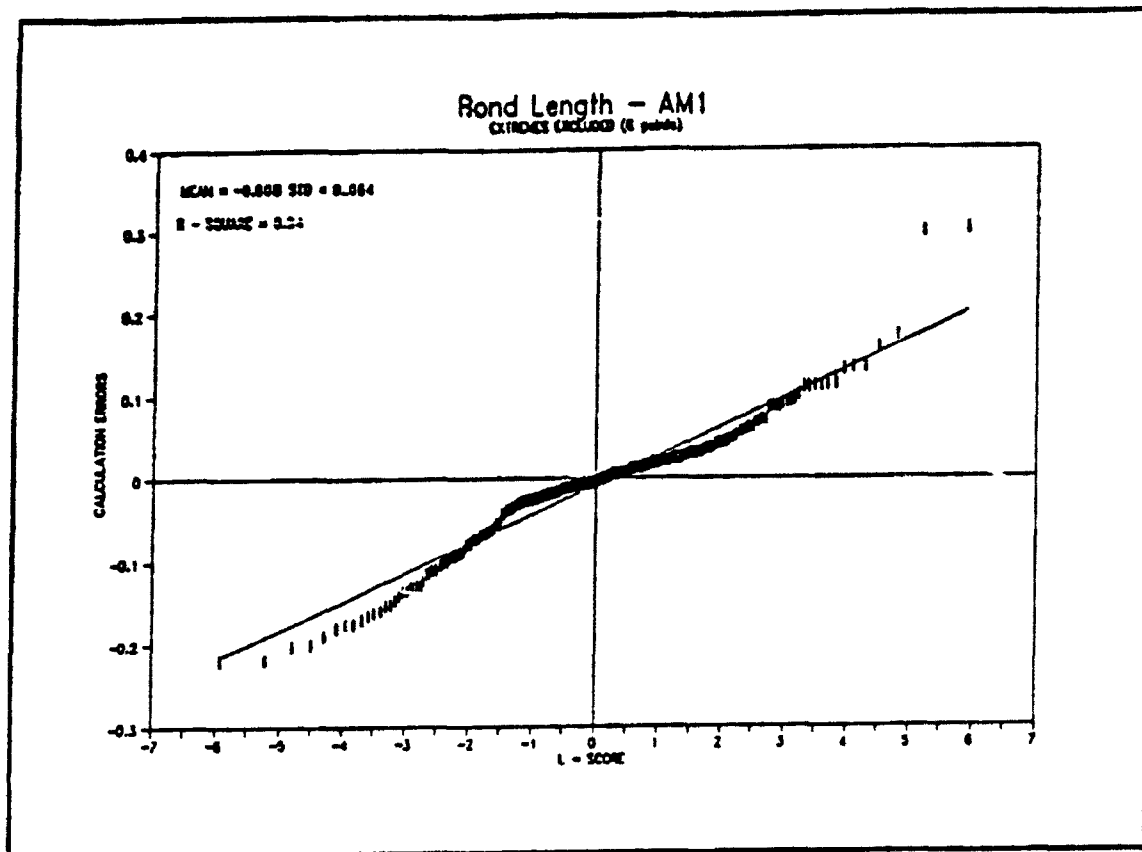


Figure 16. Test for Normal Distribution of Calculation Errors  
- Bond Length, AM1 - Extremes Excluded

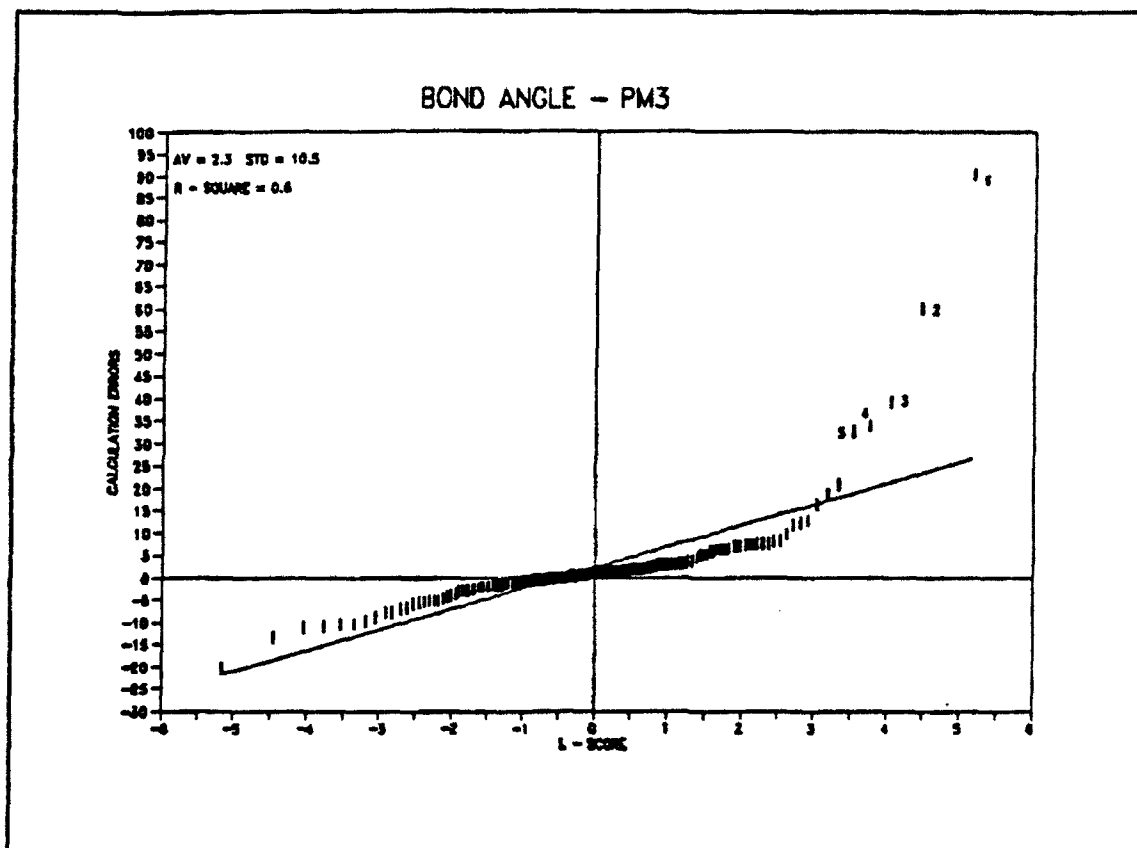


Figure 17. Test for Normal Distribution of Calculation Errors  
- Bond Angle, PM3

Table 8. Compound with Largest Deviation of Calculated Error  
of Bond Angle Using the PM3 Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	H <sub>4</sub> N <sub>2</sub>	HN-NH	90.0	90.3
2	H <sub>2</sub> O <sub>2</sub>	HO-OH	119.8	60.2
3	H <sub>2</sub> F <sub>2</sub>	H'FH	108.0	39.0
4	F <sub>3</sub> Br <sub>2</sub>	F'BrF	86.2	33.8
5	F <sub>3</sub> Cl	FClF'	87.5	32.5

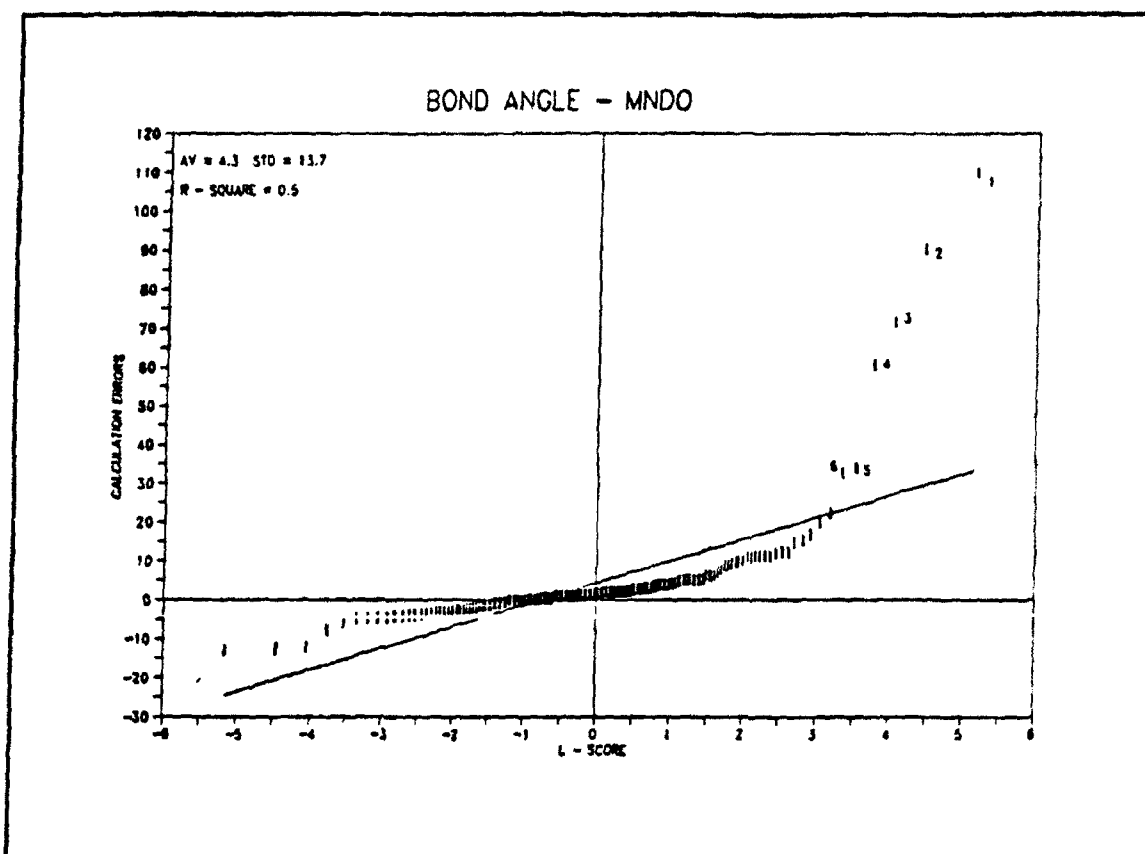


Figure 18. Test for Normal Distribution of Calculation Errors  
- Bond Angle, MNDO

Table 9. Compound with Largest Deviation of Calculated Error  
of Bond Angle Using the MNDO Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	1,4 Pentadiene	C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	-4.3	110
2	H <sub>4</sub> N <sub>2</sub>	HN-NH	90.0	90.2
3	H <sub>2</sub> F <sub>2</sub>	H'FH	108.0	71.5
4	H <sub>2</sub> O <sub>2</sub>	HO-OH	119.8	60.5
5	F <sub>3</sub> Br <sub>2</sub>	F'BrF	86.2	33.8
6	F <sub>3</sub> Cl	FClF'	87.5	32.5

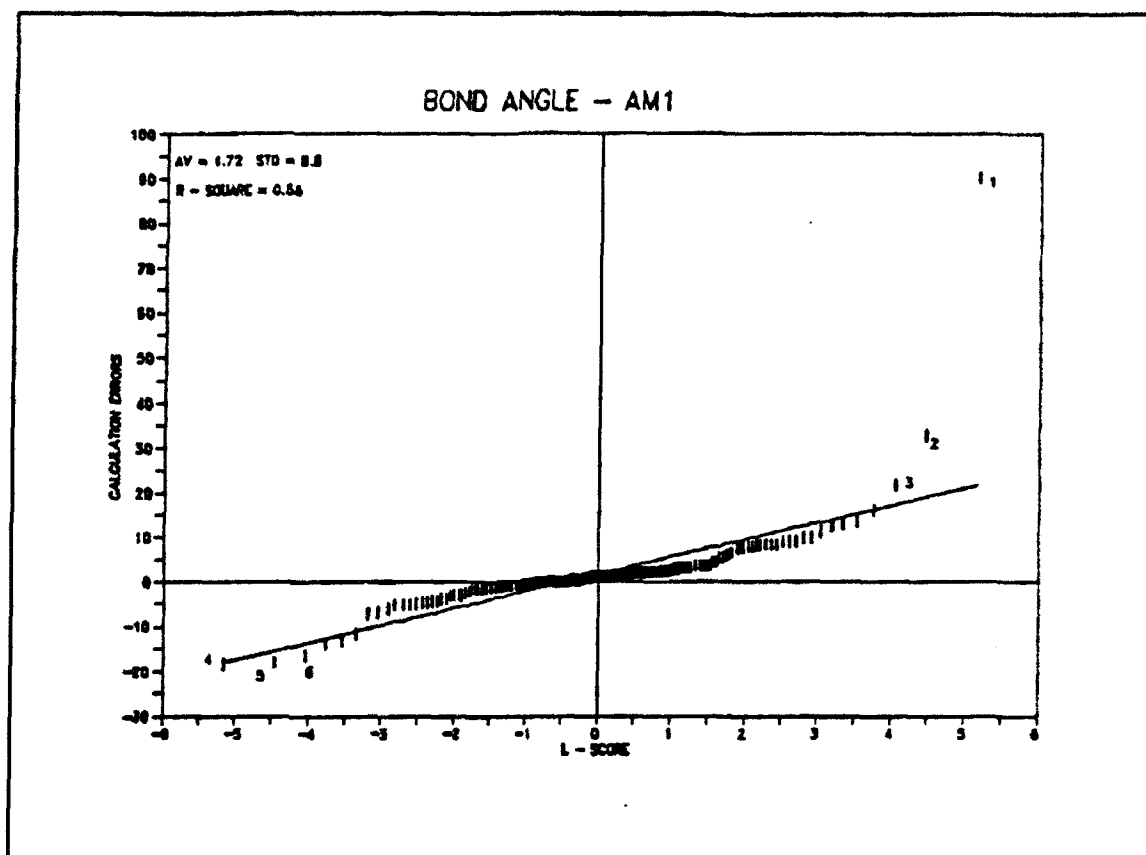


Figure 19. Test for Normal Distribution of Calculation Errors  
- Bond Angle, AM1

Table 10. Compound with Largest Deviation of Calculated Error  
of Bond Angle Using the AM1 Method

#	CHEMICAL	BOND	EXPERIMENTAL	CALC. ERRORS
1	H <sub>4</sub> N <sub>2</sub>	HN-NH	90.0	90.2
2	F <sub>3</sub> Cl	FC1F'	87.5	32.5
3	H <sub>2</sub> S <sub>2</sub>	HS-SH	90.5	21.7
4	SF <sub>4</sub>	FSF	101.6	-18.6
5	H <sub>2</sub> F <sub>2</sub>	H'FH	108	-18.1
6	SO <sub>2</sub>	OSO	119.5	-16.6



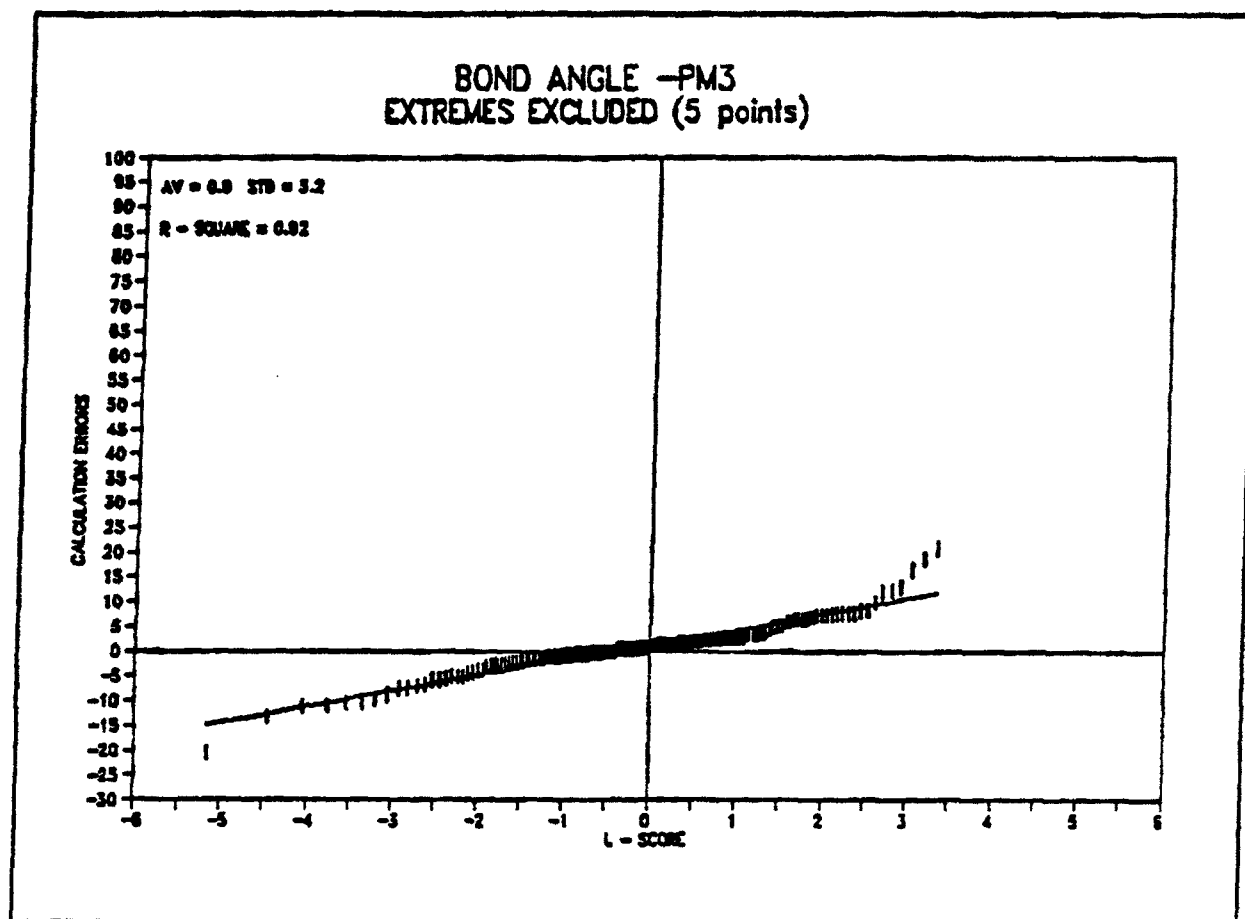


Figure 20. Test for Normal Distribution of Calculation Errors  
- Bond Angle, PM3 - Extremes Excluded

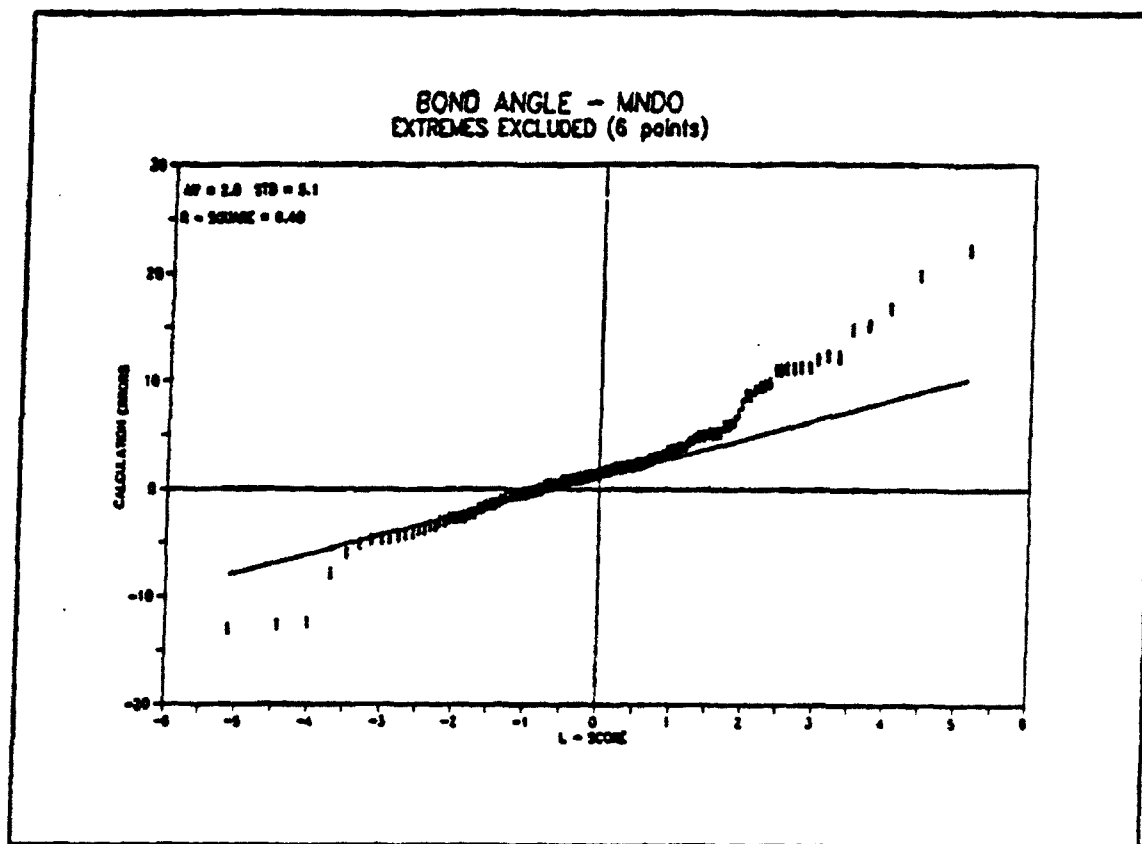


Figure 21. Test for Normal Distribution of Calculation Errors  
- Bond Angle, MNDO - Extremes Excluded

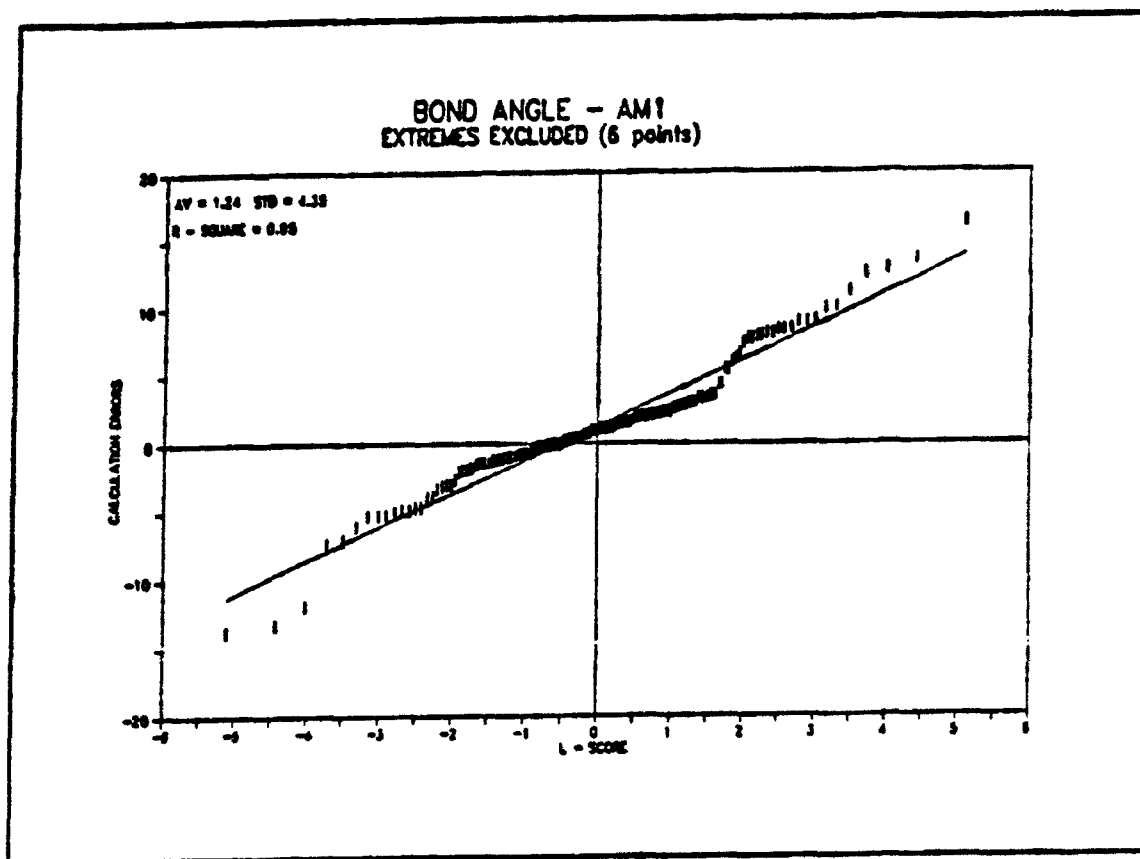


Figure 22. Test for Normal Distribution of Calculation Errors  
- Bond Angle, AM1 - Extremes Excluded

Table 11. Recommended Computational Methods for the  
Different Physical/Chemical Properties

COMPUTED PROPERTY	RECOMMENDED METHOD	BIAS	$\sigma$
HEAT OF FORMATION	PM3	0.44Kcal/mole	12.3Kcal/mole
IONIZATION POTENTIAL	MNDO	0.68 ev	0.70 ev
DIPOLE MOMENT	AM1	0.014 deby	0.49 deby
BOND LENGTH	AM1	0.08 Å	0.064 Å
BOND ANGLE	AM1	1.24°	4.35°